



[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

Search: The ACM Digital Library The Guide

"natural language" + "information retrieval" + "advances" + "

SEARCH

THE ACM DIGITAL LIBRARY

[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used [natural language](#) [information retrieval](#)
[advances](#) [text processing](#) [DARPA terms](#) [Tysons Corner](#)

Found 10,007 of 150,885

Sort results by relevance Save results to a Binder
Display results expanded form Search Tips
 Open results in a new window

Try an [Advanced Search](#)
Try this search in [The ACM Guide](#)

Results 1 - 20 of 200

Result page: **1** [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale

1 [Natural language information retrieval in digital libraries](#)

Tomek Strzalkowski, Jose Perez-Carballo, Mihnea Marinescu

April 1996 **Proceedings of the first ACM international conference on Digital libraries**

Full text available: [pdf\(1.03 MB\)](#) Additional Information: [full citation](#), [references](#), [index terms](#)

2 [Information retrieval using robust natural language processing](#)

Tomek Strzalkowski, Barbara Vauthey

June 1992 **Proceedings of the 30th conference on Association for Computational Linguistics**

Full text available: [pdf\(772.67 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
 [Publisher Site](#)

We developed a prototype information retrieval system which uses advanced natural language processing techniques to enhance the effectiveness of traditional key-word based document retrieval. The backbone of our system is a statistical retrieval engine which performs automated indexing of documents, then search and ranking in response to user queries. This core architecture is augmented with advanced natural language processing tools which are both robust and efficient. In early experiments, the ...

3 [Technique for automatically correcting words in text](#)

Karen Kukich

December 1992 **ACM Computing Surveys (CSUR)**, Volume 24 Issue 4

Full text available: [pdf\(6.23 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

Research aimed at correcting words in text has focused on three progressively more difficult problems: (1) nonword error detection; (2) isolated-word error correction; and (3) context-dependent word correction. In response to the first problem, efficient pattern-matching and n-gram analysis techniques have been developed for detecting strings that do not appear in a given word list. In response to the second problem, a variety of general and application-specific spelling cor ...

Keywords: n-gram analysis, Optical Character Recognition (OCR), context-dependent spelling correction, grammar checking, natural-language-processing models, neural net classifiers, spell checking, spelling error detection, spelling error patterns, statistical-

language models, word recognition and correction

4 Information retrieval on the web

Mei Kobayashi, Koichi Takeda

June 2000 **ACM Computing Surveys (CSUR)**, Volume 32 Issue 2

Full text available:  [pdf\(213.89 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

In this paper we review studies of the growth of the Internet and technologies that are useful for information search and retrieval on the Web. We present data on the Internet from several different sources, e.g., current as well as projected number of users, hosts, and Web sites. Although numerical figures vary, overall trends cited by the sources are consistent and point to exponential growth in the past and in the coming decade. Hence it is not surprising that about 85% of Internet user ...

Keywords: Internet, World Wide Web, clustering, indexing, information retrieval, knowledge management, search engine

5 Special issue on word sense disambiguation: Introduction to the special issue on word sense disambiguation: the state of the art

Nancy Ide, Jean Véronis

March 1998 **Computational Linguistics**, Volume 24 Issue 1

Full text available:   [pdf\(3.44 MB\)](#) [Publisher Site](#) Additional Information: [full citation](#), [references](#), [citations](#)

6 Special issue on using large corpora: II: Lexical semantic techniques for corpus analysis

James Pustejovsky, Peter Anick, Sabine Bergler

June 1993 **Computational Linguistics**, Volume 19 Issue 2

Full text available:   [pdf\(1.90 MB\)](#) [Publisher Site](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

In this paper we outline a research program for computational linguistics, making extensive use of text corpora. We demonstrate how a semantic framework for lexical knowledge can suggest richer relationships among words in text beyond that of simple co-occurrence. The work suggests how linguistic phenomena such as metonymy and polysemy might be exploitable for semantic tagging of lexical items. Unlike with purely statistical collocational analyses, the framework of a semantic theory allows the a ...

7 Information extraction as a basis for high-precision text classification

Ellen Riloff, Wendy Lehnert

July 1994 **ACM Transactions on Information Systems (TOIS)**, Volume 12 Issue 3

Full text available:  [pdf\(2.79 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

We describe an approach to text classification that represents a compromise between traditional word-based techniques and in-depth natural language processing. Our approach uses a natural language processing task called "information extraction" as a basis for high-precision text classification. We present three algorithms that use varying amounts of extracted information to classify texts. The relevancy signatures algorithm uses linguistic phrases; the a ...

Keywords: information extraction, text classification

8 Technical correspondence: Workshop on the evaluation of natural language processing systems 

Martha Palmer, Tim Finin

September 1990 **Computational Linguistics**, Volume 16 Issue 3

Full text available:  pdf(701.03 KB)

 Publisher Site

Additional Information: [full citation](#), [references](#), [citations](#)

9 TextTiling: segmenting text into multi-paragraph subtopic passages 

Marti A. Hearst

March 1997 **Computational Linguistics**, Volume 23 Issue 1

Full text available:  pdf(2.46 MB) 

 Publisher Site

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

TextTiling is a technique for subdividing texts into multi-paragraph units that represent passages, or subtopics. The discourse cues for identifying major subtopic shifts are patterns of lexical co-occurrence and distribution. The algorithm is fully implemented and is shown to produce segmentation that corresponds well to human judgments of the subtopic boundaries of 12 texts. Multi-paragraph subtopic segmentation should be useful for many text analysis tasks, including information retrieval and ...

10 Support concept-based multimedia information retrieval: a knowledge management approach 

Bin Zhu, Marshall Ramsey, Hsinchun Chen, Rosie V. Hauck, Tobun D. Ng, Bruce Schatz

January 1999 **Proceeding of the 20th international conference on Information Systems**

Full text available:  pdf(1.56 MB)

Additional Information: [full citation](#), [references](#), [index terms](#)

11 SCISOR: extracting information from on-line news 

P. S. Jacobs, Lisa F. Rau

November 1990 **Communications of the ACM**, Volume 33 Issue 11

Full text available:  pdf(1.35 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

The future of natural language text processing is examined in the SCISOR prototype. Drawing on artificial intelligence techniques, and applying them to financial news items, this powerful tool illustrates some of the future benefits of natural language analysis through a combination of bottom-up and top-down processing.

12 The impact on retrieval effectiveness of skewed frequency distributions 

Mark Sanderson, C. J. Van Rijsbergen

October 1999 **ACM Transactions on Information Systems (TOIS)**, Volume 17 Issue 4

Full text available:  pdf(145.10 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

We present an analysis of word senses that provides a fresh insight into the impact of word ambiguity on retrieval effectiveness with potential broader implications for other processes of information retrieval. Using a methodology of forming artificially ambiguous words, known as pseudowords, and through reference to other researchers' work, the analysis illustrates

that the distribution of the frequency of occurrence of the senses of a word plays a strong role in ambiguity's impact of effe ...

Keywords: pseudowords, word sense ambiguity, word sense disambiguation

13 Training a selection function for extraction

Chin-Yew Lin

November 1999 **Proceedings of the eighth international conference on Information and knowledge management**

Full text available:  pdf(2.38 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

In this paper we compare performance of several heuristics in generating informative generic/query-oriented extracts for newspaper articles in order to learn how topic prominence affects the performance of each heuristic. We study how different query types can affect the performance of each heuristic and discuss the possibility of using machine learning algorithms to automatically learn good combination functions to combine several heuristics. We also briefly describe the design, implementa ...

Keywords: automated text summarization, summary evaluation, topic extraction

14 Special issue on using large corpora: I: Introduction to the special issue on computational linguistics using large corpora

Kenneth W. Church, Robert L. Mercer

March 1993 **Computational Linguistics**, Volume 19 Issue 1

Full text available:

 pdf(1.80 MB) 

Additional Information: [full citation](#), [references](#), [citations](#)

[Publisher Site](#)

15 Creating segmented databases from free text for text retrieval

Lisa F. Rau, Paul S. Jacobs

September 1991 **Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available:

 pdf(904.82 KB) 

Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

16 The FINITE STRING newsletter: Site reports

Computational Linguistics Staff

April 1986 **Computational Linguistics**, Volume 12 Issue 2

Full text available:

 pdf(1.65 MB) 

Additional Information: [full citation](#)

[Publisher Site](#)

17 Special issue on using large corpora: I: Retrieving collocations from text: Xtract

Frank Smadja

March 1993 **Computational Linguistics**, Volume 19 Issue 1

Full text available:

 pdf(2.41 MB) 

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

[Publisher Site](#)

Natural languages are full of collocations, recurrent combinations of words that co-occur

more often than expected by chance and that correspond to arbitrary word usages. Recent work in lexicography indicates that collocations are pervasive in English; apparently, they are common in all types of writing, including both technical and nontechnical genres. Several approaches have been proposed to retrieve various types of collocations from the analysis of large samples of textual data. These techni ...

18 Retrieval performance in Ferret a conceptual information retrieval system

Michael L. Mauldin

September 1991 **Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available:  pdf(831.33 KB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)



19 Text and information extraction: Joining statistics with NLP for text categorization

Paul S. Jacobs

March 1992 **Proceedings of the third conference on Applied natural language processing**

Full text available:  pdf(908.36 KB)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)



Automatic news categorization systems have produced high accuracy, consistency, and flexibility using some natural language processing techniques. These knowledge-based categorization methods are more powerful and accurate than statistical techniques.

However, the phrasal pre-processing and pattern matching methods that seem to work for categorization have the disadvantage of requiring a fair amount of knowledge-encoding by human beings. In addition, they work much better at certain tasks, such ...

20 Automatic adaptation of proper noun dictionaries through cooperation of machine learning and probabilistic methods

Georgios Petasis, Alessandro Cucchiarelli, Paola Velardi, Georgios Palioras, Vangelis Karkaletsis, Constantine D. Spyropoulos

July 2000 **Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available:  pdf(756.53 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)



The recognition of Proper Nouns (PNs) is considered an important task in the area of Information Retrieval and Extraction. However the high performance of most existing PN classifiers heavily depends upon the availability of large dictionaries of domain-specific Proper Nouns, and a certain amount of manual work for rule writing or manual tagging. Though it is not a heavy requirement to rely on some existing PN dictionary (often these resources are available on the web), its coverage of a d ...

Keywords: *information extraction, machine learning and IR, natural language processing for IR, text data mining*

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2005 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)